

15K  
50K multiple pages  
use for...

- Objective
  - Who is going to submit
  - What message
- 24 hrs. Before Interview

How many Unix flavors do you support

additional servers

- Csr. references
- platforms

Send the prices -

Email users

Further 500 c/s

<FACT SHEET>

WAIS server

available + shipping now

## How WAIS fits into the Current Internet Environment

- Dec 6<sup>th</sup> available

- Beta users for last 4 weeks

### The Internet

The Internet consists of over 25,000 interconnected networks internationally. User growth has been explosive - an estimated 25 million people using email over the Internet worldwide and new users growing by more than one million a month. While a substantial portion of the users have traditionally come from the education and government sectors, businesses now account for the most rapid growth in newly registered networks.

### Finding Information on the Internet

People primarily use the Internet as a means to communicate (through electronic mail) and as a way of accessing information contained in various locations on the Internet. Software programs which let users browse, or "surf the net" are useful for seeing what is available on the Internet. It is difficult to know what information is out there since the Internet is a distributed network and there can be no central authority with the ability to create a comprehensive index of information.

- 500 - Hammer on Features -
- ① - Commercial over free WAIS server 0.3
  - ② Revised Feature list for V2.0
  - ③ References EB DJ go through it
  - ④ Platform list
  - ⑤ Price list

### The World Wide Web (WWW) and Hypertext

WAIS as a core information retrieval technology  
Core WAIS technology 1985-1986

The WWW project was started by CERN (the European Laboratory for Particle Physics) as a distributed hypertext program used over the Internet. In a hypertext document you can "point and click" on highlighted items which point to text in another location in the document. With hypertext, the highlighted items (text or graphics), can point to a wide range of media -- text, images, sound, etc. The WWW allows for hypertext browsers which can link one document to other documents in different locations on the Internet. The browser programs read and fetch documents from this "web" of existing locations on the Internet.

### Query Report -

What did the server do for you

- How end user question
- How often did things appear
- How many doc in db
- What version are you running

All through WAIS get process

Integrate to show a single search box

tool for building the forms pages

Support protocol stack of Z39.50 (NISO) ANSI, ISO

Conquest, Verity - license protocol stack from us -

freeware  
Can weave together for multi  
5-10 mb

Database Searching -

5-10 Gigabytes +  
getting larger

Incremental addition

Version 2 of Z39.50

What's important

~~or~~ oriented for database access  
www for display purposes

V1  $\leftrightarrow$  V2 Support the newer clients

Required to have a WAIS server at every location  
don't need a WAIS server -

Index on machine running the WAIS server

Distributed co. across databases over enterprise wide network

V1 of protocol may leave some documents out  
librarians protocol Extended to include  
MARC records - card catalog

HTTP Put-get protocol See this document w/URL  
add doc. through HTTP

FTP  
SMTP  
NNTP

Put/Get

SQL  
Similar

SQL  
writers

V2 as standard plan  
multimedia  
natural language searches  
large numbers of data bases  
around the world

end user oriented

Z39.50 different style

structure & passes data structure  
describes what it wants in detail

Words  
how many  
headline, URL, score, date, length  
types that doc is available in

Ask for documents

Search + retrieval  
Structure protocol - describes what it is  
looking for

1992 V2 certified  
Librarian community  
growth is being  
standardized this  
Berkeley  
L.D.C.  
U.C.

• Large database feature

• 239.50 Version 2

- better searching  
- break up db into multiple segments -  
search them as if they were one  
multiple discs testing on 80 Gb db

SQL  
script

General hook

easy to use API

External Unix programs to find documents  
for searching ~~search~~ index locally

or whatever  
it takes to get  
out of  
the db,

Gopher was originally developed in April 1991 at the University of Minnesota. The Internet Gopher client/server provides a distributed information delivery system. While providing a delivery vehicle for local information, Gopher facilitates access to other Gopher and information servers throughout the world. The Unix gopher server has the capability to be queried by WWW (World Wide Web) clients (either using built in gopher querying or using native http querying).

Veronica: Very Easy Rodent-Oriented Net-wide Index to Computerized Archives.

Veronica offers a keyword search of most gopher-server menu titles in the entire gopher web. As archie is to ftp archives, veronica is to gopherspace. A veronica search produces a menu of gopher items, each of which is a direct pointer to a gopher data source. Because veronica is accessed through a gopher client, it is easy to use, and gives access to all types of data supported by the gopher protocol.

The dominant user interface of choice on the Internet

--Internet, what it is, who uses it, explosive growth, impact on business sector

-Access to the Internet through client software - people need to find their way around this huge mass of locations, databases: people use email to communicate, but people need to browse to find different locations- surf the net, see what is out there.

-The WWW developed as a browsing program- brief history of CERN, original use - then further developed by NCSA.

-Difference between the WWW - how it works and the client programs. (Mosaic, Netscape)

-What Mosaic/Netscape can do, what they \_can't\_ do. They can't quickly search databases and retrieve useful information.

-What WAIS does- it is the power behind the WEB for searching databases on the Internet.

"WAIS unlocks the content behind the WEB"

Should we talk about Gopher- where should we bring up Gopher?

Large scale publishing, sort  
in

WAIS Gate

gateways

Script for accessing db  
db admin of original system  
must understand original database

WWW internet critical at this time in the market  
Search or browse database back-end

100 most wanted  
some web browser  
weights  
Creating the search  
also pointers to  
the files  
build the  
items -  
fielded search - relevance feedback  
data ranges

CGP 'int. info' cylob  
EB

Metadata code  
tools to help  
in creation

- where it came from

TAGS attached to  
each document

What the document  
is about

Can codes  
industry codes

Engine behind web

in original formats SGML  
text doc.

Available easily through  
web gateway or API

fielded search  
allows ~~for~~ metadata codes  
- dates  
- authors

index creation

• Create an indexed version of document

different

Oracle or Sybase SQL call or  
proprietary dump it in a particular format  
existing parser or custom parser

Mitch Wagner / Open Systems Today

↳ announcing

Adding large data collections into the web

native product publishers

Freeware 1991

Professional  
version

Web Servers - what happens now

Content available, but not navigable

- Can't get what they want w/ hypertext

Answering

Professional uses

How to  
know when a  
Database is  
powered by  
WAIS -  
Instead of hypertext  
links,

Currently there  
are a number of  
web browser mss. including  
Dino 2

gained  
which ~~has~~  
~~feeding~~ ~~was~~  
~~popular~~ ~~was~~  
~~popular~~

## WWW Browsers

<sup>In Sept of 1993</sup>  
The National Center for Supercomputing Applications (NCSA) released the PC/Windows and the Macintosh versions of the "Mosaic" WWW browser in September of 1993. Since that time web browsers have gained widespread popularity on the Internet. Some of the web browsers currently in use include WinWeb, Lynx, Spyglass Mosaic, Internet Works, and Netscape. These browsers provide ease of use and the sensory appeal of integrated access to different types of media (graphics, text, sound, animations). Central to the usefulness of the web browsers or "clients" is their ability to work in conjunction with other existing Internet search tools.

## WAIS and WWW as complimentary technologies

WAIS searching is an integral part of the resources available through the WWW. The WWW opens the door to a wider part of the Internet than was previously available through Gopher, a popular browsing program which predates web browsers. Using the web clients for browsing, "surfing the net", does not take full advantage of the resources available on the WWW. Placing WAIS technology behind the web enables the user to take a step beyond browsing to unlock the content. WAIS technology addresses the issue of organizing the resources to make them readily available and easy to search.

Two important methods for navigating the web are: browsing & searching. The <sup>web</sup> ~~mentioned~~ browsers solve much of the problem of browsing. WAIS provides the technology behind the searching.

Newspapers have the ability to search for specific ~~items~~ topics and also browse through the stories -

WAIS Delete get rid of content without re-index the whole database

Incremental creates a small new database & merge it into the ~~new~~ database

DV - multi-database as if it was one no down time

### Xerox Corp page

No mechanism for searching for  
Spec Specks on the web

publishers side -

don't want to recede everything in HTML

Automated approach to make available to the web

### Existing tools

- re-index whole database  
hand translate
- Natural language search tools  
boolean or not elaborate  
word weighting librarians +  
professional searching  
too hard to use

Created database of pointers  
new development

- become a professional system  
smooth + easily
- large databases
- tuned parameters
- incremental addition -

< word weighting  
returns ranked list

Netmanage - ?

Galaxy - JS using free version

boolean  
bad word weighting } one or two words

Freeware - works well for small database  
as you scale up need more sophisticated -  
Commercial version

Traditional  
Local Searching

Index your HTML pages  
islands on the internet  
too many pages to visit  
find data over many servers